

Anahtar Sözcük Arama için Skor Düzgeleme

Score Normalization for Keyword Search

Leda Sarı, Murat Saraçlar
Elektrik ve Elektronik Mühendisliği Bölümü
Boğaziçi Üniversitesi
İstanbul, Türkiye
{leda.sari,murat.saraclar}@boun.edu.tr

Özetçe —Bu çalışmada, ses verisinin posteriorgram gösterimiini kullanarak simgesel dizin tabanlı anahtar sözcük arama (ASA) yapılmıştır. Arama sonuçlarına her sorgu için skorların toplamı bir olacak şekilde düzgeleme ya da sorguya özgü eşikleme gibi farklı skor düzgeleme yöntemleri uygulanmıştır. Bu yöntemlerin önerilen ASA sisteminin başarısına etkisi incelenmiştir. Sonuçlar otomatik konuşma tanıma (OKT) örüelerinden elde edilen dizine dayalı bir ASA sistemiyle birleştirilmiş ve düzgelemenin birleştirilmiş sisteme katmasına bakılmıştır. Başarım ölçütü olarak maksimum terim ağırlıklı değer (MTAD) kullanılmış, birleştirilmiş sonuçlarda özellikle OKT dağarcığının dışında (DD) kalan sorgular için MTAD'nın artırıldığı gösterilmiştir.

Anahtar Kelimeler—anahtar sözcük arama, dizinleme, skor düzgeleme.

Abstract—In this work, keyword search (KWS) is based on a symbolic index that uses posteriorgram representation of the speech data. For each query, sum-to-one normalization or keyword specific thresholding is applied to the search results. The effect of these methods on the proposed KWS system is investigated. Results are combined with a KWS system that is based on an index generated from automatic speech recognition (ASR) lattices and the effect of score normalization on the performance of the combined system is observed. Maximum term weighted value (MTWV) is used as the performance measure. It is shown that the MTWV is increased in the combined system especially for the out-of-vocabulary (OOV) keywords that are not covered by the ASR lexicon.

Keywords—keywords search, indexing, score normalization.

I. GİRİŞ

Anahtar sözcük aramanın (ASA) amacı yazılı olarak verilen sorguların yazılandırılmamış ses verilerinin içindeki yerlerinin saptanmasıdır. Bu sorgu ya da “anahtar sözcük” bir veya daha fazla kelimedenden oluşabilir. Mevcut ASA sistemlerinin çoğu otomatik konuşma tanıma (OKT) çıktılarına metin geri

This study uses the IARPA Babel Program base period language collection release babel105b-v0.4, and is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Defense U.S. Army Research Laboratory (DoD/ARL) contract number W911NF-12-C-0012. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoD/ARL, or the U.S. Government.

getirimi yöntemlerini uygular. Bu amaçla OKT örüelerinden bir dizin oluşturulur ve sorgu bu dizinde aranır [1], [2]. OKT tabanlı ASA sistemlerindeki sorunlardan biri dağarcık dışı (DD) sorguların aranmasıdır çünkü dizinde OKT sözlüğünde bulunmayan bu sözcüklere ait bilgi bulunmaz. Bu problemin çözümünde kullanılan yöntemlerden biri sözcükler yerine seslik, hece gibi sözcük altı birimlerden arama dizini oluşturmaya ve DD sorgunun bu sözcük altı birimler cinsinden ifade edilmesine dayanır [1], [3]. Diğer bir yöntem ise DD sözcükleri aramak yerine onlara akustik olarak benzeyen dağarcık içi (Dİ) sözcükleri aramaktır. Benzer sözcükler aranacak DD sözlüğün bir karışıklık modeli (KM) uygulanarak genişletilmesiyle elde edilir [4], [5]. Arama sonuçları her bir anahtar sözcüğün hangi dosyaların hangi zaman aralıklarında geçtiği bilgisini ve her bir bulguya ait bir güvenilirlik skoru içerir. Sistem başarısının ölçülmesinde ise terim ağırlıklı değer (TAD) kullanılır. TAD bulguların skorlarına bağlı olduğundan ve skorlar da sorgunun uzunluğu, eğitim verisinde görülmeye sıklığı gibi etkenlere bağlı olduğundan skorların düzgelenme şekli TAD'yi etkiler [6], [7].

ASA sistemlerine benzeyen diğer bir uygulama da konuşulan terimleri örnek-ile-sorguya saptamadır. Bu yöntemler genellikle sorgu ile aranacak verideki örüntüler eşleştirme yöntemlerine dayanır. Bu çalışmalarda veri gösteriminde genellikle posteriorgramlar kullanılır [8], [9].

Bu çalışmada ASA sisteminde, OKT örüelerinden oluşturulan dizine benzer ağırlıklı sonlu durum dönüştürücüsü (ASDD) yapısında bir dizin elde edilmiş ve arama bu dizin üzerinde yapılmıştır. Bu dizin aranacak verinin posteriorgramının simgesel bir gösterimi kullanılarak oluşturulmuştur [10]. DD sorguların aranmasında KM ile genişletme kullanılmıştır. ASA sonucunda elde edilen ham skorlara farklı düzgelemeler uygulanıp TAD ile ölçülen sistem başarısını nasıl etkileiği incelenmiştir. Ayrıca, önerilen ASA sisteminin sonuçlarının mevcut OKT tabanlı ASA sistemiyle birleştirilmesiyle özellikle DD sorgular için TAD'nın artırılması hedeflenmiştir.

Bildirinin ikinci bölümünde simgesel dizin tabanlı ASA sistemi tanıtılacak, üçüncü bölümde ASA sisteminin başarım ölçütü ve skor düzgeleme yöntemleri anlatılacak, dördüncü bölümde deney düzeneği ve deney sonuçları sunulup beşinci bölümde sonuçlar özetlenerek bildiri sonlanacaktır.

II. SİMGESİSEL DİZİNLE ANAHTAR SÖZCÜK ARAMA

Posteriorgram, verinin zamana ve sınıfa dayalı iki boyutlu bir gösterimidir. Her bir zaman aralığına karşılık gelen vektör,

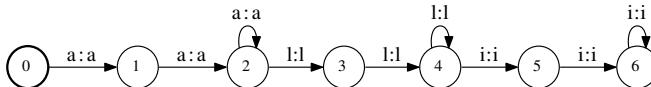
sınıfların o zaman aralığındaki sonsal olasılık dağılımlarını gösterir [8]. Bu vektörler, OKT'de kullanılan derin yapay sinir ağlarının (DYSA) çıkış katından elde edilir. Ses kayıtlarındaki her çerçeve için, o verinin öznitelikleri DYSA'dan geçirilerek, verinin posteriorogramı elde edilir.

A. Simgesel Dizinle Anahtar Sözcük Arama

Aranacak veri kümesi sözcelere ayrılmış her birinin posteriorogramı elde edildikten sonra bir kod tablosu kullanılarak bir simge dizisine dönüştürülür. Her bir sözce bir simge dizisine çevrildikten sonra OKT örülerinden elde edilen dizine benzer şekilde ASDD biçiminde bir dizin oluşturulur [2]. Bu ASDD'de durumlar arasındaki geçişlerin girdi etiketleri simgeleri, çıktı etiketleri ise sözce numarasını gösterir, ağırlıkları ise simgelerin başlangıç ve bitiş zamanı bilgilerini içerir. ASDD biçiminde bir dizin kullanılması sorgu olarak herhangi bir ASDD'nin kullanılmasına olanak tanır [2], [11]. Örneğin, sorgu genişletme yapıldığında doğrusal yapıda bir ASDD yerine alternatif simge dizilerini de içeren yapıların aranmasına izin verir.

Simge dizilerinin çıkarılmasında kullanılan kod tablosu ortalama topaklama gibi denetimsiz öğrenmeyle çıkarılabileceği gibi eğitim verisi kullanılarak da çıkarılabilir. Eğitim verisi posteriorogramı ve bu verinin sesçik hizalaması kullanılarak her sesçiye karşılık gelen posteriorogram vektörlerinin ortalaması alınır. Bu durumda kod tablosu her sesçikle eşleşen bir kod vektörü içerir.

Aramaya hazırlık aşamasında yazılı olarak verilen sorgular Bölüm II-B'de anlatılacağı gibi bir simge dizisine çevrilir. Bu simgelerden doğrusal yapıya sahip bir ASDD yerine simgelerin en az belirli sayıda tekrar etmesini gerektiren döngü yapıları eklenir. Örneğin, Şekil 1'de gösterilen "Ali" sorgusunun ASDD'si için bulunan eşleşmede her simbolün en az iki kez tekrar etmesi gereklidir.



Şekil 1: Döngüler içeren sorgu ASDD'si

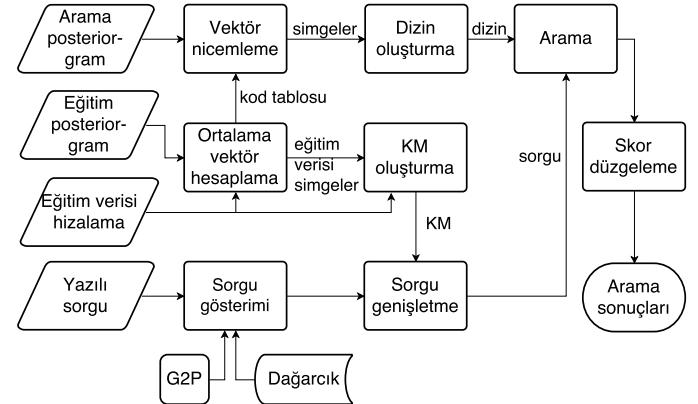
Bu sorgu daha sonra bir KM ile genişletilerek alternatif simge dizilerinin aranması ve böylece sonuç sayısının artırılması amaçlanır. KM aynı zamanda nicemleme hatalarını da telafi etmekte kullanılır. Burada kullanılan KM ASDD yapısındadır ve hangi simgelerin birbirile karişıklarını modeller. Karışma ağırlıkları ise girdi-çıktı simbol çiftinin yer değiştirmeye olasılığının negatif logaritmasıdır. Eğitim verisinin posteriorogramı dizinlemede kullanılan kod tablosuyla nicemlenir ve bu verinin simgesel gösterimi elde edilir. Bu dizi eğitim verisinin sesçik hizalamasıyla karşılaştırılarak bir karışıklık matrisi elde edilir. Bu matris K ile gösterildiğinde $K(i, j)$ elemanı j simgesinin i yerine kaç kez kullanıldığını gösterir. Bu matris düzgelenerek karışma olasılıkları belirlenir. Düzgelenmiş matrisin belirli bir eşigi aşmış elemanları kullanılarak KM oluşturulur.

ASDD tabanlı arama aşamasında genişletilmiş sorgu ve dizin ASDD'lerinin bileşkesi alınır. Elde edilen yeni ASDD, arama skoru, başlangıç ve bitiş zamanı bilgilerini içerir. Bu

ASDD'nin tüm yollarını sonuç olarak döndürmek yerine en yüksek skora sahip yollar tutulur (n -best işlemi). Sorgu, KM ve dizin S , KM ve D ile gösterildiğinde ve ASDD bileşke işlemi o ifade edildiğinde, arama sonuçları (1)'deki gibi özetlenebilir:

$$\text{Sonuç} = n\text{-best}(S \circ KM \circ D) \quad (1)$$

Simgesel dizin tabanlı sistemin akış çizeneği de Şekil 2'de gösterilmiştir.



Şekil 2: Simgesel dizin tabanlı sistemin akış çizeneği

B. Sorgunun Simgesel Gösterimi

Sorgunun aramaya hazırlanmasında öncelikle yazılı olarak verilen sorgu, dağarcık ve harften sese dönüştüren bir model (G2P) kullanılarak sesçik dizisine çevrilir. Dizinlemede kullanılan kod tablosu her sesçiye karşılık bir vektör içeriysa bu sesçik dizileri doğrudan kullanılabilir. Diğer bir yöntemse sesçik dizisinden bir posteriorogram elde edip bu posteriorogramı kod tablosuya nicemleyip simge dizisine çevirmektir. Posteriorogramın modellenmesinde ise iki farklı yöntem kullanılır.

İlk yöntemde posteriorogram vektörleri sadece ilgili sesçiye karşılık gelen eleman 1 diğer elemanlar 0 olacak şekilde modellenir. İkinci yöntemde ise posteriorogram vektörleri eğitim verisinden çıkarılan her bir sesçiye karşılık gelen ortalama vektör kullanılarak oluşturulur.

III. ASA SİSTEMİNİN DEĞERLENDİRİLMESİ

Denklem (1)'den elde edilen sonuçlar belirli bir θ eşik değerinin üzerindeyse bu sonuçlar bulgu olarak döndürülür. Bu bölümde, bulguların değerlendirilmesinde kullanılan terim ağırlıklı değer (TAD) ve bu ölçütü iyileştirmek için başvurulan skor düzgeleme yöntemlerinden bahsedilecektir.

A. Terim Ağırlıklı Değer

TAD, belirli bir θ değerinin üzerinde kalan arama sonuçları kullanılarak (2)'deki gibi hesaplanır [12].

$$\begin{aligned} TAD(\theta) &= 1 - \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} P_{\text{miss}}(q, \theta) + \beta P_{\text{FA}}(q, \theta) \\ &= 1 - \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} 1 - \frac{N_{\text{cor}}(q, \theta)}{N_{\text{true}}(q)} + \beta \frac{N_{\text{spur}}(q, \theta)}{N_{\text{NT}}(q)} \end{aligned} \quad (2)$$

Bu ifadede Q tüm sorguların kümesini, $|Q|$, Q 'deki eleman sayısını, $P_{\text{miss}}(q, \theta)$ ve $P_{\text{FA}}(q, \theta)$ q sorgusunun θ eşik değeri kullanıldığındaki kaçırma ve yanlış alarm verme olasılıklarını göstermektedir. İkinci satırdaki N_{true} , N_{cor} , N_{spur} , N_{NT} sırasıyla q sorgusunun kayıt bütüncesinde (*corpus*) geçme sayısını, arama sonucunda doğru bulunan hipotezlerinin sayısını, yanlış alarm sayısını ve toplam süre ile N_{true} 'nin farkını, dolayısıyla olabilecek tüm yanlış alarm verme sayısını göstermektedir.

TAD'ın alabileceği en yüksek değer 1'dir ve tüm sorguların doğru şekilde bulunmasına karşılık gelir. Hiçbir sonuç döndürmeyen sistemler için TAD 0'dır ve TAD eksiz değerler alabilir. Olabilecek tüm eşik değerleri için elde edilebilecek en yüksek TAD de maksimum TAD (MTAD) olarak adlandırılır [13].

B. Skor Düzgeleme

Sistem başarımı θ 'ya bağlı olduğundan skorlara düzgeleme uygulamak başarımı etkiler. Ayrıca farklı düzgeleme işlemleri, farklı sayıda sonuca sahip sorgular arasında bir denge sağlamak ya da aranan sorgunun uzunluğundan bağımsız skorlar elde etmeyeceğini kullanır [14].

Düzgeleme yöntemlerinden biri her bir sorgu için o sorguya ait skorların toplamını (TB) 1 yapmaktadır. Bir q sorgusuna ait h bulgusunun skoru $p_{q,h}$ ve bu sorguya ait toplam bulguların kümesi H_q olsun. Bu durumda düzgelenmiş skor $\bar{p}_{q,h}$, (3)'teki gibi bulunur.

$$\bar{p}_{q,h} = \frac{p_{q,h}}{\sum_{h' \in H_q} p_{q,h'}} \quad (3)$$

Bir bulguya ait skorların dinamik erimi geniş olduğunda TB düzgeleme sırasında γ -düzgelemesi de yapılabilir:

$$\bar{p}_{q,h} = \frac{p_{q,h}^\gamma}{\sum_{h' \in H_q} p_{q,h'}^\gamma} \quad (4)$$

Bu düzgelemede $\gamma = 1$ alınırsa TB düzgeleme yapılmış olur, $\gamma = 0$ olduğunda ise bir sorguya ait tüm sonuçlara eşit skor atanmış olur.

Sorguya özgü eşikleme (SE) yönteminde ise kaçırma ve yanlış alarm vermeden kaynaklanacak zarar olasılıkları eşitlenerek her sorguya özgü bir eşik değeri $\theta(q)$ bulunur.

$$\theta(q) = \frac{\alpha \sum_h p_{q,h}}{T/\beta + \frac{\beta-1}{\beta} \alpha \sum_h p_{q,h}} \quad (5)$$

Bu denkleme $\alpha \sum_h p_{q,h}$, tahmini N_{true} değeri olarak kullanılmıştır. TAD'de tek bir eşik değeri kullanılacağı için, skorlar ait olduğu sorgunun eşik değeri kullanılarak (6)'ya göre düzgelenir [15].

$$\bar{p}_{q,h} = (p_{q,h})^{\log(\theta)/\log(\theta(q))} \quad (6)$$

IV. DENEYLER

Deneylerde IARPA Babel programındaki Türkçe (babel105b-v0.4) konuşma verileri üzerinde çalışılmıştır. Bu veriler karşılıklı telefon konuşmalarını içerir. ASA deneylerinde 88'i DD olan 307 sorgu, 10 saatlik veri içerisinde aranmıştır.

Posteriorogramların çıkarılmasında Kaldi araç takımının Babel deney düzeneğindeki [16] DYSA kullanılmıştır. Bu DYSA p-norm doğrusalsızlığı kullanır.

Sistemlerin tek başlarına başarımının yanı sıra sonuçlar, mevcut OKT tabanlı bir ASA sistemiyle birleştirilerek özellikle DD sorgular için başarım artırılması hedeflenmiştir. Temel sistem olarak Kaldi araç takımının Babel deney düzeneği kullanılmış [16], farklı sistemlerin sonuçları CombMNZ yöntemiyle birleştirilmiştir [14]. Temel sistemde skor düzgeleme için SE kullanılmış ve $\alpha = 1$ alınmıştır.

Simgesel dizin tabanlı sistemde her bir sesçiye karşılık bir sesçik posteriorogram vektörü olacak şekilde bir kod tablosu kullandığımızda ve KM'yi eğitim verisini kullanarak oluşturduğumuzda Tablo I'deki sonuçlar elde edilmiştir. Bu tabloda

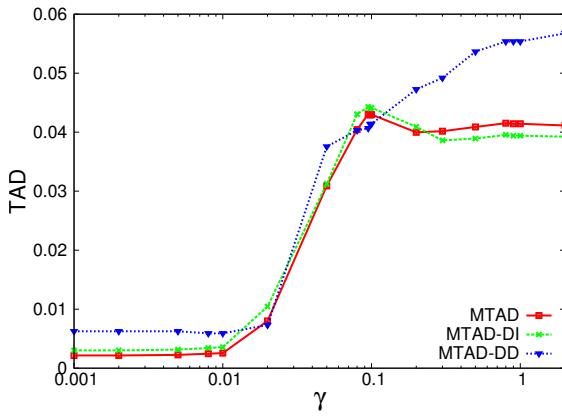
Yöntem	Parametre	MTAD		
		Tümü	Dİ	DD
TB	-	0.0414	0.0394	0.0554
α	2.0	0.0410	0.0386	0.0536
α	1.5	0.0410	0.0387	0.0530
α	0.2	0.0396	0.0390	0.0558
γ	0.095	0.0431	0.0443	0.0407
γ	2.0	0.0411	0.0392	0.0568

Tablo I: Düzgelemeye bağlı olarak simgesel dizin tabanlı ASA sisteminin TAD'ları

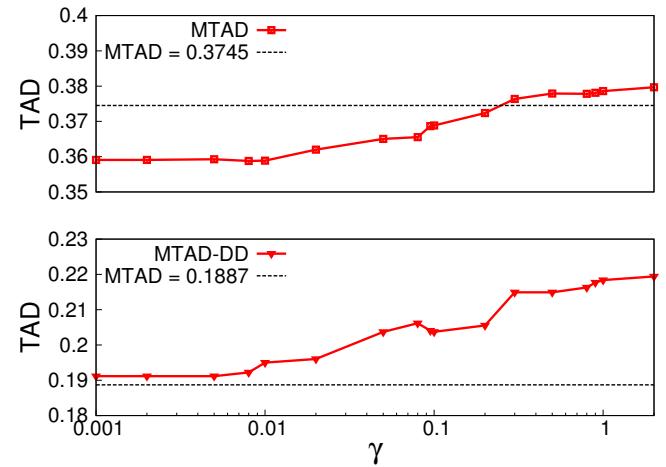
ilk sütunda hangi düzgeleme yönteminin kullanıldığı, ikinci sütunda ise düzgelemede kullanılan değişkenin (γ ya da α) değeri verilmiştir. MTAD, tüm sorgular, Dİ ve DD sorgular üzerinden ayrı ayrı hesaplanmıştır. Şekil 3'te MTAD'ların γ düzgelemede değişkenin değerine göre değişimini gösterilmiştir. Burada tüm sorgular üzerinden hesaplanan en iyi MTAD $\gamma = 0.095$ iken elde edilmiş, DD sorgular üzerinden hesaplanan MTAD (MTAD-DD)'de artan bir eğri gözlemlenmiştir. Skorlar 0-1 aralığında olduğundan, γ büyük seçildiğinde (4)'teki pay ve payda 0 olur dolayısıyla sonuçların süzgeçlenmesi gereklidir. Bu nedenle $\gamma > 2$ olduğunda MTAD değerleri düşmeye başlamıştır ve MTAD-DD, $\gamma = 2.0$ olduğunda en yüksek değerini almıştır. Bu değerler ($\gamma = 0.095$ ve $\gamma = 2.0$) için hesaplanan MTAD'ler Tablo I'de verilmiştir. Tabloya göre, tüm sorgular üzerinden düşünüldüğünde TB düzgelemesi ($\gamma = 1.0$) de $\gamma = 2.0$ durumuna benzer bir sonuç vermiştir.

Tablo II'de OKT örüsunü dayalı ASA sistemi (Örü) ve düzgeleme uygulanmış simgesel dizin tabanlı sistemin birleştirilmiş sonuçları gösterilmiştir. 'DD' ile gösterilen deneylerde önerilen sistemin sadece DD sonuçları 'Örü'yle birleştirilmiştir. Birleştirmede tüm deneylerde MTAD-DD artsa da en iyi sonuçlara TB düzgeleme yapıldığında ulaşılmıştır. Bu durumda MTAD-DD'de Örü'ye göre %16 artış görülmüştür.

Şekil 4'te γ düzgelemesi uygulanmış simgesel dizin tabanlı ASA sisteminin DD sonuçları OKT örüsü tabanlı ASA sistemiyle birleştirildiğinde elde edilen TAD'nın γ 'ya göre değişimini gösterilmiştir. Önerilen sistemin Dİ sonuçları birleşime katılmadığından Dİ sorgular için MTAD sabit ve 0.45'tir. Şekilde görüldüğü gibi tüm γ değerleri için MTAD-DD, yatay çizgiyle gösterilen temel sistemin DD sonucundan (0.1887) daha yüksektir. En yüksek değerler ise γ değeri 1'e



Şekil 3: Simgesel dizin tabanlı sisteme TAD'ının γ 'ya göre değişimi



Şekil 4: Birleştirilmiş sisteme TAD'ının γ 'ya göre değişimi

Sistem	Parametre	MTAD		
		Tümü	Dİ	DD
Örü	-	0.3745	0.4500	0.1887
Örü + TB	-	0.3650	0.4321	0.2184
Örü + TB _{DD}	-	0.3786	0.4500	0.2184
Örü + α	2.0	0.3583	0.4265	0.2101
Örü + α_{DD}	2.0	0.3748	0.4500	0.2101
Örü + α	1.5	0.3587	0.4260	0.2094
Örü + α_{DD}	1.5	0.3769	0.4500	0.2094
Örü + α	0.2	0.3580	0.4229	0.2143
Örü + α_{DD}	0.2	0.3773	0.4500	0.2143
Örü + γ	0.095	0.3556	0.4251	0.2041
Örü + γ_{oov}	0.095	0.3687	0.4500	0.2041

Tablo II: Sistem birleştirmesi sonunda elde edilen MTAD, KTAD and ETAD, Örü+X: İki sistemin tüm sonuçları, Örü+X_{DD}: Örü (Dİ ve DD) ve önerilen ASA sisteminin DD sonuçlarının birleşimi

yakın olduğunda elde edilmiştir. Tüm sorgular için MTAD de γ , 1'e yakın değerler aldığımda temel sisteme göre artmıştır.

V. SONUÇLAR

Bu çalışmada ses verisinin posteriorgram gösteriminden simgesel bir dizin elde edilmiş, yazılı olarak verilen sözcükler bu dizinde aranmıştır. Bu ASA sisteminin döndürdüğü sonuçlara skor düzgeleme uygulanmış, hem bu sistemde hem de sistemin OKT örusü tabanlı mevcut bir ASA sistemiyle birleştirildiğinde MTAD ile ölçülen sistem başarısını nasıl etkilediği gösterilmiştir. Önerilen ASA sisteminin sonuçları düzgelendikten sonra OKT örusü tabanlı mevcut bir ASA sistemiyle birleştirildiğinde özellikle DD sorgularda MTAD'yi mevcut sisteme göre %16 artırdığı gözlenmiştir. TB düzgeleme, sistem birleştirmede SE ve γ düzgelemelerine göre daha iyi sonuç verdiğimden ve seçilmesi gereken bir parametre içermemişinden ASA sistemlerinde tercih edilebilir.

KAYNAKÇA

- [1] M. Saraclar and R. Sproat, "Lattice-based search for spoken utterance retrieval," in *HLT-NAACL*, pp. 129–136, 2004.
- [2] D. Can and M. Saraclar, "Lattice indexing for spoken term detection," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2338–2347, 2011.
- [3] D. Can, E. Cooper, A. Sethy, C. White, B. Ramabhadran, and M. Saraclar, "Effect of pronunciations on oov queries in spoken term detection," in *Proceedings of ICASSP*, pp. 3957–3960, 2009.
- [4] P. Karanasou, L. Burget, D. Vergyri, M. Akbacak, and A. Mandal, "Discriminatively trained phoneme confusion model for keyword spotting," in *Proceedings of Interspeech*, p. 2434–2437, 2012.
- [5] M. Saraclar, A. Sethy, B. Ramabhadran, L. Mangu, J. Cui, X. Cui, B. Kingsbury, and J. Mamou, "An empirical study of confusion modeling in keyword search for low resource languages," in *IEEE ASRU*, pp. 464–469, 2013.
- [6] V. Soto, L. Mangu, A. Rosenberg, and J. Hirschberg, "A comparison of multiple methods for rescoring keyword search lists for low resource languages," *Proceedings of Interspeech*, pp. 2464–2468, 2014.
- [7] Y. Wang and F. Metze, "An in-depth comparison of keyword specific thresholding and sum-to-one score normalization," *Proceedings of Interspeech*, pp. 2474–2478, 2014.
- [8] T. J. Hazen, W. Shen, and C. White, "Query-by-example spoken term detection using phonetic posteriorgram templates," in *Proc. ASRU*, pp. 421–426, 2009.
- [9] Y. Zhang, *Unsupervised speech processing with applications to query-by-example spoken term detection*. PhD thesis, MIT, 2013.
- [10] L. Sari, B. Gündoğdu, and M. Saraclar, "Fusion of Ivcsr and posteriorgram based keyword search," in *Proceedings of Interspeech*, pp. 824–828, 2015.
- [11] C. Parada, A. Sethy, and B. Ramabhadran, "Query-by-example spoken term detection for oov terms," in *Proc. ASRU*, pp. 404–409, 2009.
- [12] J. G. Fiscus, J. Ajot, J. S. Garofolo, and G. Doddington, "Results of the 2006 spoken term detection evaluation," in *Proceedings of ACM SIGIR Workshop on Searching Spontaneous Conversational*, pp. 51–55, 2007.
- [13] "Open KWS14 Keyword Search Evaluation Plan," <http://www.nist.gov/itl/iad/mig/upload/KWS14-evalplan-v11.pdf>.
- [14] J. Mamou, J. Cui, X. Cui, M. J. Gales, B. Kingsbury, K. Knill, L. Mangu, *et al.*, "System combination and score normalization for spoken term detection," in *Proceedings of ICASSP*, pp. 8272–8276, 2013.
- [15] D. Karakos, R. Schwartz, S. Tsakalidis, L. Zhang, S. Ranjan, T. Tim Ng, Hsiao, *et al.*, "Score normalization and system combination for improved keyword spotting," in *Proc. ASRU*, pp. 210–215, IEEE, 2013.
- [16] "Kaldi Babel recipe," <https://github.com/kaldi-asr/kaldi/tree/master/egs/babel/s5c>.